Newcastle University

MAS1403

Quantitative Methods for Business Management

Semester 2, 2011—12

Lecturer: Dr. Lee Fawcett

Contact details

My name: Dr. Lee Fawcett

Office: Room 2.07 Herschel Building

Phone: 0191 222 7228

Email: lee.fawcett@ncl.ac.uk

www: www.mas.ncl.ac.uk/∼nlf8

My involvement with this course

MAS1403 *used* to be taught by the Business School (pre–2005)

I have been **module leader** for this course since Maths & Stats took it over in 2005

- I have taught both semesters
- I have been heavily involved with setting up "innovative" teaching tools for this course
  - CBAs (2007)
  - Video material (see webpage)
  - Revision DVD (this will be given out at Easter)
  - Computerised exam

My involvement with this course

- I know that most of you probably **resent** having to take this course...
- ...however, by the end of the module many students **appreciate** the need for a course like this in their degree programme
- And **most do quite well** (average mark last year – 67!)
- I *think* I am very **approachable** – please feel free to drop by my office if you are stuck with anything!
- Don't just sit on any problems – the material in semester 2 is quite hard, and builds week–by–week. If you are stuck, come to see me **as soon as you can**!
- Make the most of **tutorials**: students have told me in the past that they are often more important that the lectures themselves!
- Oh – I will be taking registers in tutorials as well!

Timetable

- **Lectures** are in the Curtis Auditorium in the Herschel Building on Mondays at 2pm (Business Management/Accounting & Finance) or 3pm (everyone else).
- **Tutorials** have changed! The new times/locations are:
  - **Group A**: Tuesday 12, Herschel LT3 (Lee)
  - **Group B**: Tuesday 3, Herschel LT3 (Malcolm)
  - **Group C**: Thursday 3, Herschel LT2 (Lee)
  - **Group D**: Thursday 4, Herschel LT2 (Lee)
  - **Group E**: Friday 9, Herschel LT2 (Malcolm)
  - **Group F**: Friday 10, Herschel LT2 (Malcolm)
  - **Group G**: Friday 1, Herschel LT3 (Lee)

  Please check the Semester 2 website for allocations. In week 7, tutorials will be replaced with computer practical sessions.

Assessment

There will be three **CBA**s

There will be one **written assignment** over the Easter holidays

There will be an **exam** at the end of Semester 2 covering material from the *entire year*!

You should refer to the **week–by–week schedule** for this course for CBA/assignment deadlines, computer practicals etc. etc.

My teaching style

- Lecture notes will be posted to the course webpage:

  `www.mas.ncl.ac.uk/~nlf8/Teaching/MAS1403/home.html`

  These will usually be available on the Thursday before the lecture the following week. It is your responsibility to print these off and **bring them to the class**
- The notes often have **gaps** in them for you to fill in during the lecture – the colour band will change to purple when it's time for you to write stuff down
- Please **ask questions** if you're stuck during lectures!
- I may get you to do calculations in lectures – always bring your **calculator**!
- Make sure you come to **tutorials**!
- Lots of **revision**!

Lecture 1

CONFIDENCE INTERVALS

Recap and introduction

Recall that data can be summarised in **two** ways:

1. Graphical summaries

- Stem–and-leaf plots;
- Bar charts;
- Histograms;
- Relative frequency histograms;
- Frequency polygons.

2. Numerical summaries

- **Measures of location**
  - (i) Sample mean;
  - (ii) Sample median;
  - (iii) Sample mode.

- **Measures of spread**
  - (i) Range;
  - (ii) Variance (and standard deviation);
  - (iii) Interquartile range.

Estimation

What does our sample tell us about the population?

- We can rarely observe the entire population, so the **population mean** and **population variance** are hardly ever known *exactly*;
- These unknown quantities are called **parameters**;
- We use Greek letters to denote them – μ for the mean, and σ^2 for the variance (and so σ for the standard deviation);
- We hope that the **sample mean** ($\bar{x}$) will be quite close to the true mean (μ);
- **But how do we know if it is?**

Example: *The Vintage Clothing Co.*

The *Vintage Clothing Co.* are a large retailer of bespoke and retro clothing.

They have 1,000 branches across the U.K., and all of their branches are open on Sundays.

However, they are considering whether or not it is worthwhile staying open on Sundays.

Table 1.1 shows the number of transactions at each of their shops on Sunday 29th January 2012.

Example: *The Vintage Clothing Co.*

282	258	399	271	343	285	247	513	171	123	168	327	430	240	410	341
290	263	446	185	330	111	243	376	139	351	311	389	546	321	393	487
264	320	217	257	349	640	97	298	393	454	363	354	360	326	199	502
293	407	362	270	344	263	290	263	50	253	345	581	229	264	304	394
499	276	412	323	310	177	248	178	409	275	278	307	495	515	232	432
339	404	371	262	336	218	274	483	211	245	316	381	432	233	223	447
202	133	356	408	224	379	197	278	235	509	171	232	429	315	326	602
242	389	219	206	393	437	306	152	294	271	230	398	346	344	379	347
305	174	291	261	214	532	335	63	100	357	190	347	208	420	322	463
389	236	445	378	255	301	308	150	289	453	464	273	211	450	222	250
320	420	357	160	372	99	316	218	248	322	145	399	433	393	403	361
261	279	369	342	168	322	304	254	99	503	303	212	105	166	257	422
346	370	235	355	65	340	420	338	568	644	164	288	319	159	324	208
268	340	305	361	319	519	293	380	286	431	402	329	363	330	612	248
446	588	304	454	164	240	293	478	540	339	245	257	222	471	469	273
277	216	555	401	380	338	212	476	77	363	140	451	329	66	217	461
522	111	119	316	116	471	142	336	277	101	518	264	226	256	539	324
333	332	404	362	202	204	341	80	333	267	439	136	343	389	244	370
372	595	314	182	470	192	555	374	368	192	225	321	435	403	316	312
192	63	407	125	253	89	70	186	491	342	122	367	106	334	161	177
454	122	286	39	361	262	316	272	285	201	191	162	229	334	278	231
644	297	398	118	246	148	478	167	337	344	395	334	255	401	504	304
192	507	41	457	405	306	282	446	195	512	252	510	557	191	321	404
377	240	441	308	346	265	375	332	580	130	353	426	95	588	332	109
263	529	172	529	315	257	481	260	297	382	438	64	226	185	369	275
190	340	337	224	363	212	371	229	175	388	332	315	389	452	266	393
219	400	378	241	616	551	359	489	314	450	645	224	320	405	182	251
240	471	293	240	184	296	617	565	206	147	169	401	140	462	389	310
323	351	187	544	387	425	353	175	378	484	205	295	413	189	559	251
213	574	579	325	246	206	419	306	471	264	270	300	278	131	561	328
281	403	256	348	183	161	444	482	338	268	313	252	179	414	444	266
203	269	450	322	459	183	212	242	144	406	401	174	605	270	487	494

Example: *The Vintage Clothing Co.*

Suppose the marketing department of *The Vintage Clothing Co.* are interested in the **average number of transactions** across all their stores on Sunday 29th January 2012.

Can we work this out *exactly*?

The answer is “yes” — we have data from every single branch!

Actually, Table 1.1 shows we have taken a **census** – every single branch has been asked to provide us with data.

So in this case, it is possible to work out the **population mean** μ:

$$\mu = \frac{282 + 258 + 399 + 271 + \ldots + 426 + 477}{1000} = 320 \text{ transactions.}$$

Example: *The Vintage Clothing Co.*

Now, let's suppose the company don't have the **time**/**resources** to take a census.

In fact, a week before the 29th January, just **five stores** are selected at random — we work out the mean number of transactions using the data from these five stores only.

Let's suppose the top left–hand block in Table 1.1 above are stores 1–100, the next block along are stores 101–200, etc.

We put the numbers 1–1000 into a bag and draw, without replacement, **5 numbers at random**.

Example: *The Vintage Clothing Co.*

Store	No. of transactions, X
637	$x_1 = 374$
327	$x_2 = 452$
849	$x_3 = 271$
666	$x_4 = 419$
680	$x_5 = 643$

Let's suppose this is the only information we have.

It is no longer possible to work out the true population mean, as we don't have information from every single shop; we can now only work out the **sample mean** $\bar{x}$:

$$\bar{x} = \frac{374 + 452 + 271 + 419 + 643}{5} = 431.8 \approx 432 \text{ transactions.}$$

Example: *The Vintage Clothing Co.*

Obviously, the marketing team are not just interested in what goes on in these five shops.

However, this is the only information they have, and so they use this information to draw conclusions about all 1,000 shops as a whole.

This is known as the process of **statistical inference** – we are trying to *infer* things about the population, based on the limited information in our sample.

Hopefully, provided we don't have a **biased sample**, $\bar{x}$ will do a good job at estimating μ. Has it done a good job here?

Example: *The Vintage Clothing Co.*

The true **population mean** is $\mu = 320$.

Our **sample mean** is $\bar{x} \approx 432$.

We have **over–estimated** the population mean by quite a bit – our estimate is 112 too high!

Perhaps we have a **biased sample** – we seem to have selected a number of stores with a higher–than–average number of transactions – especially store 680!

Example: *The Vintage Clothing Co.*

This begs the question:

"*How accurate can our sample mean be in estimating the population mean?*"

Let's take another random sample by drawing another five numbers from the bag:

Store	No. of transactions, X
558	$x_1 = 253$
428	$x_2 = 446$
903	$x_3 = 251$
364	$x_4 = 256$
14	$x_5 = 185$

This gives

$$\bar{x} = \frac{253 + 446 + 251 + 256 + 185}{5} = 278.2 \approx 278 \text{ transactions.}$$

This sample mean is much closer to the population mean, but still not *very* close.

Also, it is quite different from the mean of the previous sample.

You could repeat this procedure yourself (filling in the table on page 6) to select three more random samples of size 5, and calculate the sample means.

In fact, consider this your first piece of **homework** to do before the tutorials this week!

Example: *The Vintage Clothing Co.*

Your random sample 1

$u =$	$store =$	$x_1 =$
$u =$	$store =$	$x_2 =$
$u =$	$store =$	$x_3 =$
$u =$	$store =$	$x_4 =$
$u =$	$store =$	$x_5 =$
		$\bar{x} =$

How close are these sample means to the correct population value $\mu = 320$ transactions?

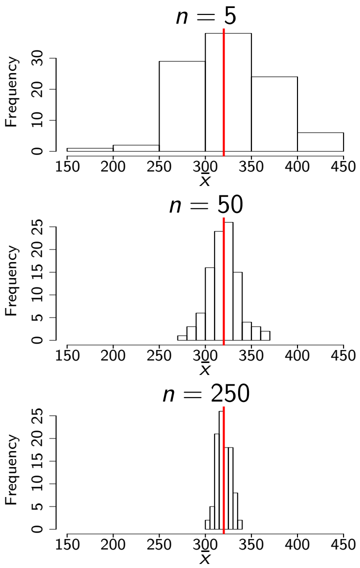
Example: *The Vintage Clothing Co.*

In fact, we could take many samples, and it's very likely that we'll get a **different value** for $\bar{x}$ each time.

It's also very *unlikely* that any of our $\bar{x}$'s will be exactly the same as the **true population mean** μ.

Figure 1.1 shows histograms of $\bar{x}$'s taken from 100 samples from our population of 1,000 stores.

Example: *The Vintage Clothing Co.*



You should notice two things:

1. The distribution of $\bar{x}$'s looks like a **Normal distribution**;
2. As we increase the sample size (n), the distribution for $\bar{x}$ gets more and more **concentrated** – around the **true population value** $\mu = 320$!

In fact, what we can see in action in this graph is known as the **Central Limit Theorem**.

This is a very powerful result in Statistics which tells us about the distribution of the sample mean $\bar{x}$. We now state this formally.

The Central Limit Theorem

Suppose $x_1, x_2, \ldots, x_n$ are a random sample from *any* population, with mean μ and variance σ^2.

If n is large, then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{approximately;}$$

if $x_1, x_2, \ldots, x_n$ come from a Normal distribution themselves, then this result holds for *any* n.

The Central Limit Theorem

In other words, if we were to take many samples of size n and:

- For each sample calculate the mean $\bar{x}$;
- Put all of our $\bar{x}$'s together and make a histogram of them;

then our histogram of $\bar{x}$'s will **always** be Normally distributed around the true population mean μ.

What's more, we also know about the **variability** of $\bar{x}$:

$$\text{var}(\bar{x}) = \frac{\sigma^2}{n}.$$

Thus, the standard deviation of $\bar{x}$ is $\sigma/\sqrt{n}$, and we call this the **standard error**.

Interval estimation

The values we calculate for sample means and variances are **point estimates**.

They are single values based on a limited sample of the whole population.

Suppose that we wish to estimate the mean μ of a population. The natural estimate for μ is the sample mean $\bar{x}$.

However, as we have seen, $\bar{x}$ is never exactly equal to μ; all we really hope is that $\bar{x}$ will be close to μ.

Interval estimation

One way of improving our inference is to construct **interval estimates**, more commonly known as **confidence intervals**.

We simply place an interval over the point estimate for μ which allows us to say (with a certain level of confidence) within what range the population mean lies.

The calculation of these intervals depends on the size of our sample (n), the level of confidence we choose, and whether or not the population variance (σ^2) is known.

Case 1: Known variance σ^2

We know from the CLT that, if our random sample is drawn from a Normal distribution, or if n is large (i.e. $n \geq 30$), then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

If we initially assume we know the population variance σ^2, we can **standardise** $\bar{x}$ as we did last semester; i.e.

$$Z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}.$$

Recall that the standard Normal distribution is $Z \sim N(0,1)$, i.e. Z has zero mean and variance (and so standard deviation) 1.

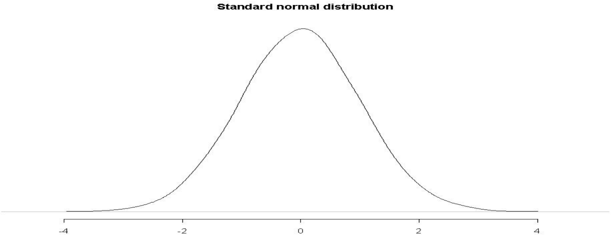
Also recall that approximately 95% of the standard normal distribution lies between -1.96 and 1.96, i.e.

Case 1: Known variance σ^2

- We know that (from tables)

$$\Pr(-1.96 < Z < 1.96) = 0.95;$$

We can think about this graphically:

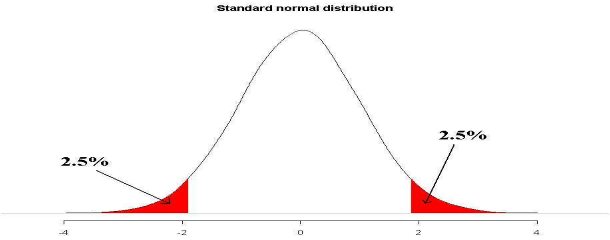


Case 1: Known variance σ^2

- We know that (from tables)

$$\Pr(-1.96 < Z < 1.96) \quad = \quad 0.95;$$

We can think about this graphically:

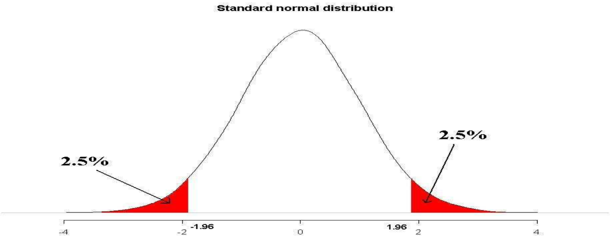


Case 1: Known variance σ^2

- We know that (from tables)

$$\Pr(-1.96 < Z < 1.96) = 0.95;$$

We can think about this graphically:

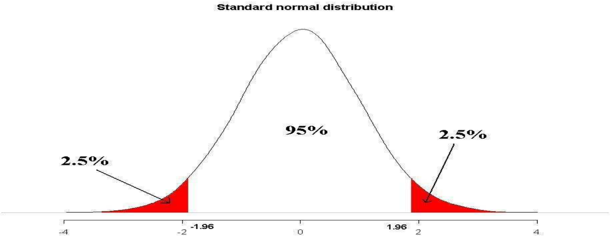


Case 1: Known variance σ^2

- We know that (from tables)

$$\Pr(-1.96 < Z < 1.96) \quad = \quad 0.95;$$

We can think about this graphically:



- Thus,

$$\Pr\left(-1.96 < \frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} < 1.96\right) \quad = \quad 0.95;$$

Case 1: Known variance σ^2

Rearranging for μ gives us an expression for the **95% confidence interval for μ**:

$$\left(\bar{x} - 1.96\sqrt{\sigma^2/n} \quad , \quad \bar{x} + 1.96\sqrt{\sigma^2/n}\right);$$

thus, we can say that the two values

- $\bar{x} - 1.96\sqrt{\sigma^2/n}$ and
- $\bar{x} + 1.96\sqrt{\sigma^2/n}$

are the **lower** and **upper bounds** (respectively) of the (95%) confidence interval.

We often write this more simply as

$$\bar{x} \pm 1.96\sqrt{\sigma^2/n}.$$

Case 1: Known variance σ^2

Going back to *The Vintage Clothing Co.* example, this means that if we were to take 100 samples and for each one calculate a 95% confidence interval, then about 95 of these confidence intervals would "capture" the true population value $\mu = 320$.

If we wanted to be "more confident" of capturing μ, then the interval needs to be wider.

We replace the value **1.96** with

- **2.58** for a 99% interval
- **1.65** for a 90% interval

If we increase our sample size, we become more certain of our estimate $\bar{x}$ and so the width of the interval **decreases**.

See a demonstration of this on page 10.

Example 2: *Geordie Sparkz*

Geordie Sparkz are an electrical company based in Newcastle producing circuitboards for large plasma televisions.

One of their machines punches tiny holes in these curcuitboards that should be 0.5mm in diameter. A sample of 30 circuitboards off the production line is inspected; the average diameter of the holes produced by this machine, for this sample, is 0.54mm.

Assuming the machine is set to ensure a standard deviation of $\sigma = 0.12$mm, calculate the 95% confidence interval for the population mean diameter of holes produced by this machine.

Do you think there is a real problem with this machine?

Example 2: *Geordie Sparkz*

We need to use the formula

$$\bar{x} \pm 1.96 \times \sqrt{\frac{\sigma^2}{n}}.$$

We have

$$n = 30, \quad \bar{x} = 0.54 \quad \text{and} \quad \sigma = 0.12.$$

This give

$$0.54 \quad \pm \quad 1.96 \times \sqrt{\frac{0.12^2}{30}}$$

$$0.54 \quad \pm \quad 1.96 \times \sqrt{\frac{0.0144}{30}}$$

$$0.54 \quad \pm \quad 1.96 \times \sqrt{0.00048}$$

$$0.54 \quad \pm \quad 0.0429$$

Example 2: *Geordie Sparkz*

This gives the 95% confidence interval as

$$(0.497\text{mm}, 0.583\text{mm}).$$

There is no real cause for concern, as the specified value (0.5mm) falls within this interval.